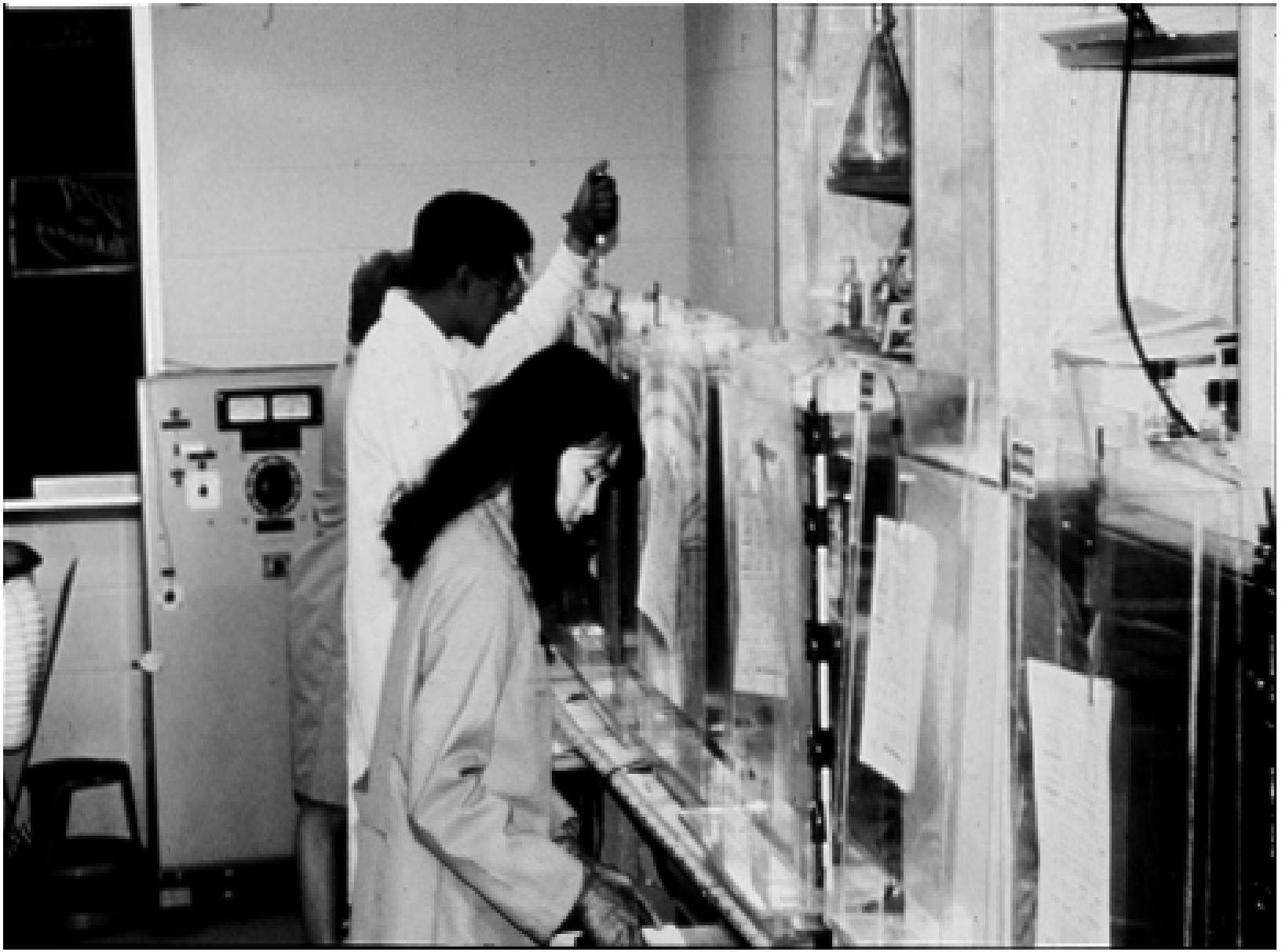
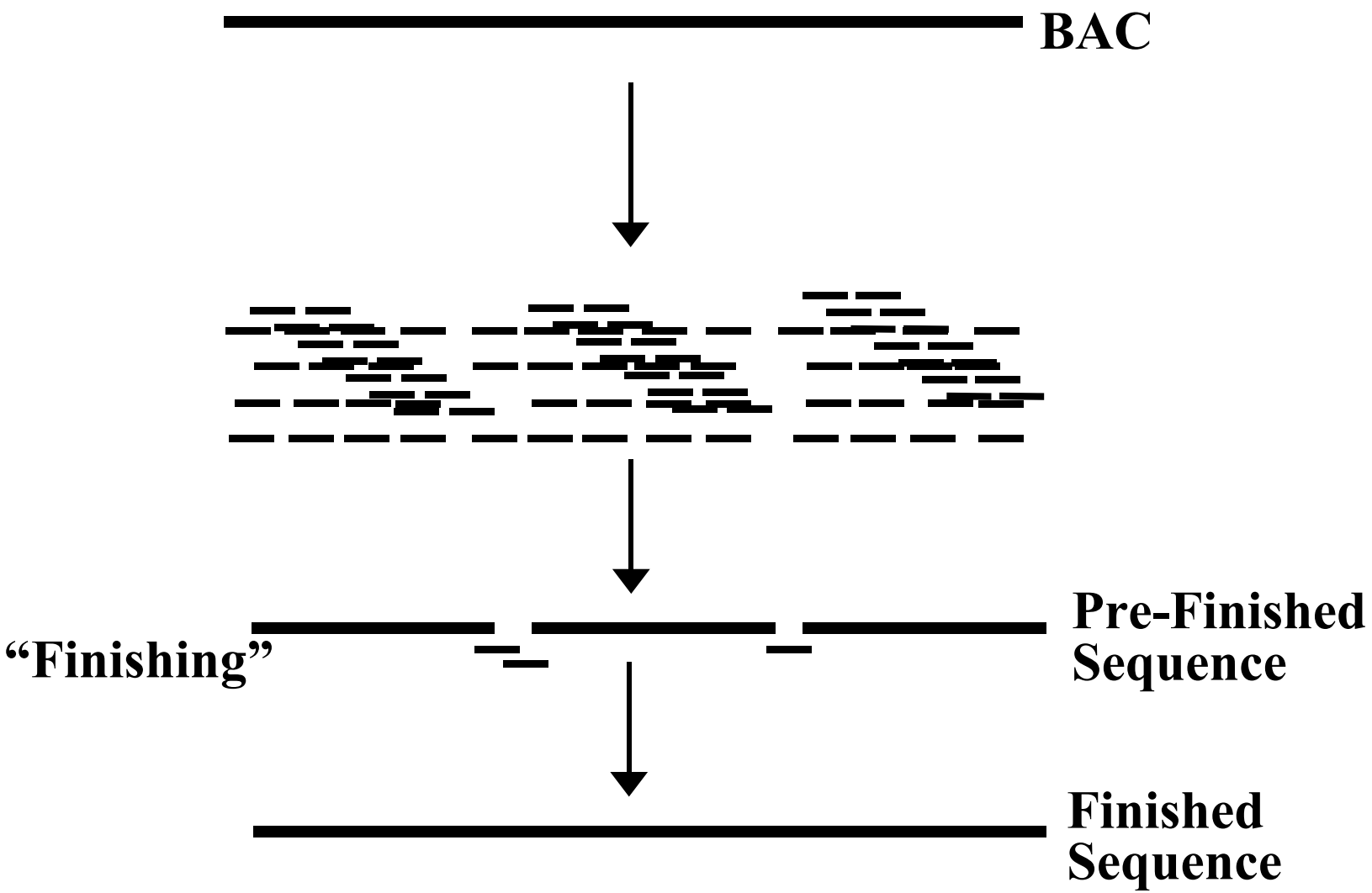
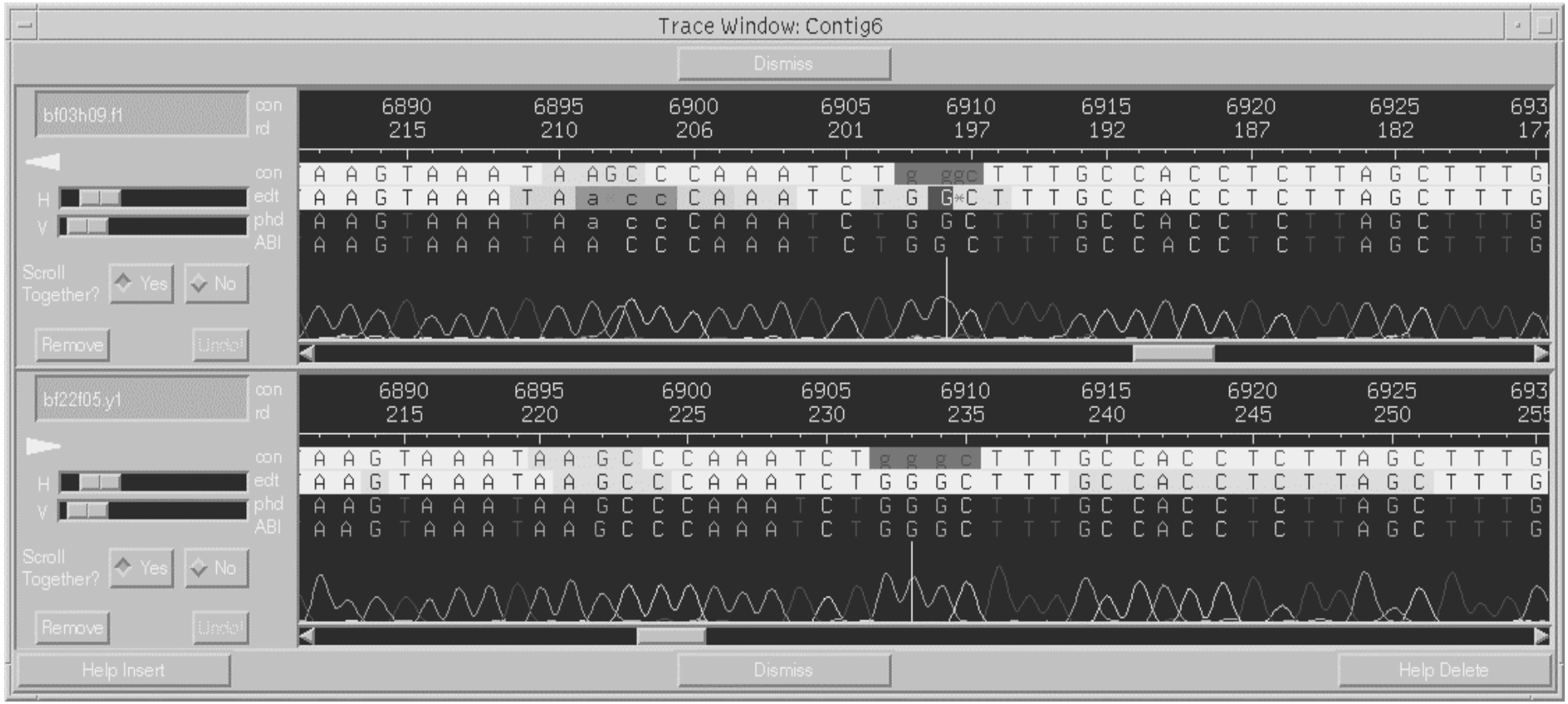
Radioactive Sequencing



Shotgun Sequencing Strategy



Resolve Ambiguities...



Why Sequence an Entire Genome?

- Provides basis for discovery of all genes
- Reveals long-range relationships between genes
- Provides the structural and control elements between genes
- Provides tools for future experiments
- Provides a framework to draw in and organize all genetic information about the organism
- The whole sequence is an archive for the future, containing all the information required to make the organism

Netscape: TIGR Microbial Database

File Edit View Go Communicator Help

Bookmarks Netsite: http://www.tigr.org/tdb/mdb/mdb.html What's Related

TIGR Microbial Database:

a listing of microbial genomes and chromosomes completed and in progress

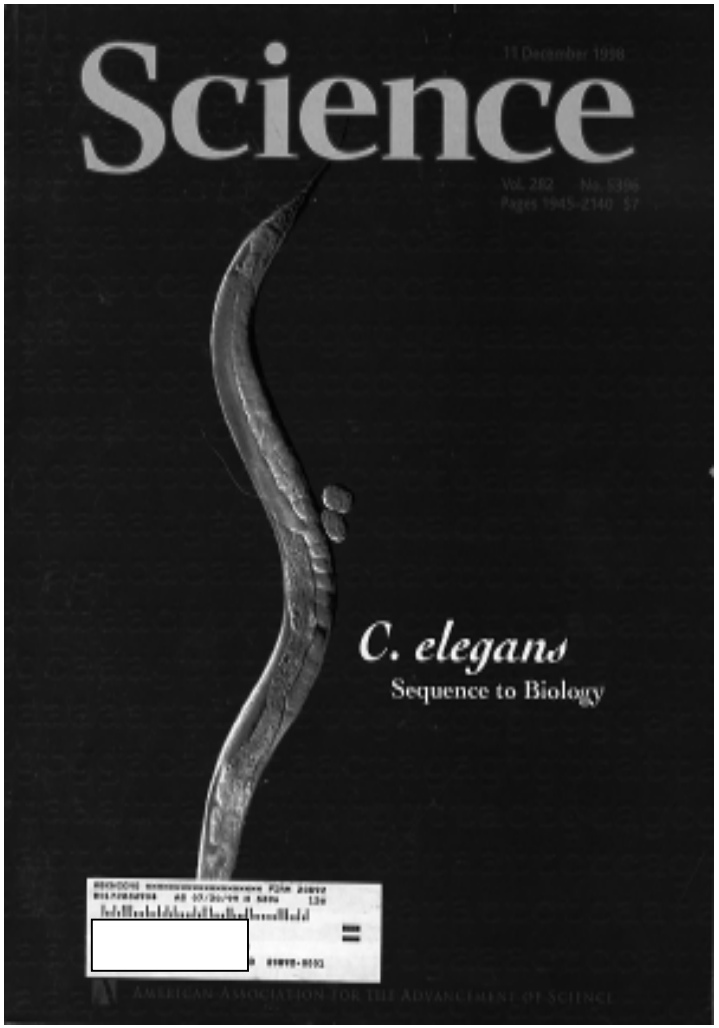
---

Published microbial genomes and chromosomes (scroll down for genomes in progress)

	Link	Genome	Strain	Domain	Size (Mb)	Institution	Funding	Publication
1		*Haemophilus influenzae* Rd	KW20	B	1.83	TIGR	TIGR	Fleischmann *et. al.*, *Science* **269**:496–512 (1995)
2		*Mycoplasma genitalium*	G–37	B	0.58	TIGR	DOE	Fraser *et. al.*, *Science* **270**:397–403 (1995)
3		*Methanococcus jannaschii*	DSM 2661	A	1.66	TIGR	DOE	Bult *et. al.*, *Science* **273**:1058–1073 (1996)
4		*Synechocystis* sp.	PCC 6803	B	3.57	Kazusa DNA Research Inst.		Kaneko *et. al.*, *DNA Res.* **3**: 109–136 (1996)
5		*Mycoplasma pneumoniae*	M129	B	0.81	Univ. of Heidelberg	DFG	Himmelreich et. al., *Nuc. Acid Res.* **24**:4420–4449 (1996)
6	SGD mips	*Saccharomyces cerevisiae*	S288C	E	13	International Consortium	EC, NHGRI, Welcome Trust, McGill U., RIKEN	Goffeau et. al., *Nature* **387** (Suppl.) 5–105 (1997)
7		*Helicobacter pylori*	26695	B	1.66	TIGR	TIGR	Tomb *et. al.*, *Nature***388**:539–547 (1997)
8		*Escherichia coli*	K–12	B	4.60	University of Wisconsin	NHGRI	Blattner *et. al.*, *Science* **277**:1453–1474 (1997)
9		*Methanobacterium*	delta H	A	1.75	Genome Therapeutics	DOE	Smith *et.al.*, *J. Bacteriology*

100%

First Complete Sequence of Multicellular Organism



Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology

The *C. elegans* Sequencing Consortium*

***Science* 282:1012-2018, 1998**

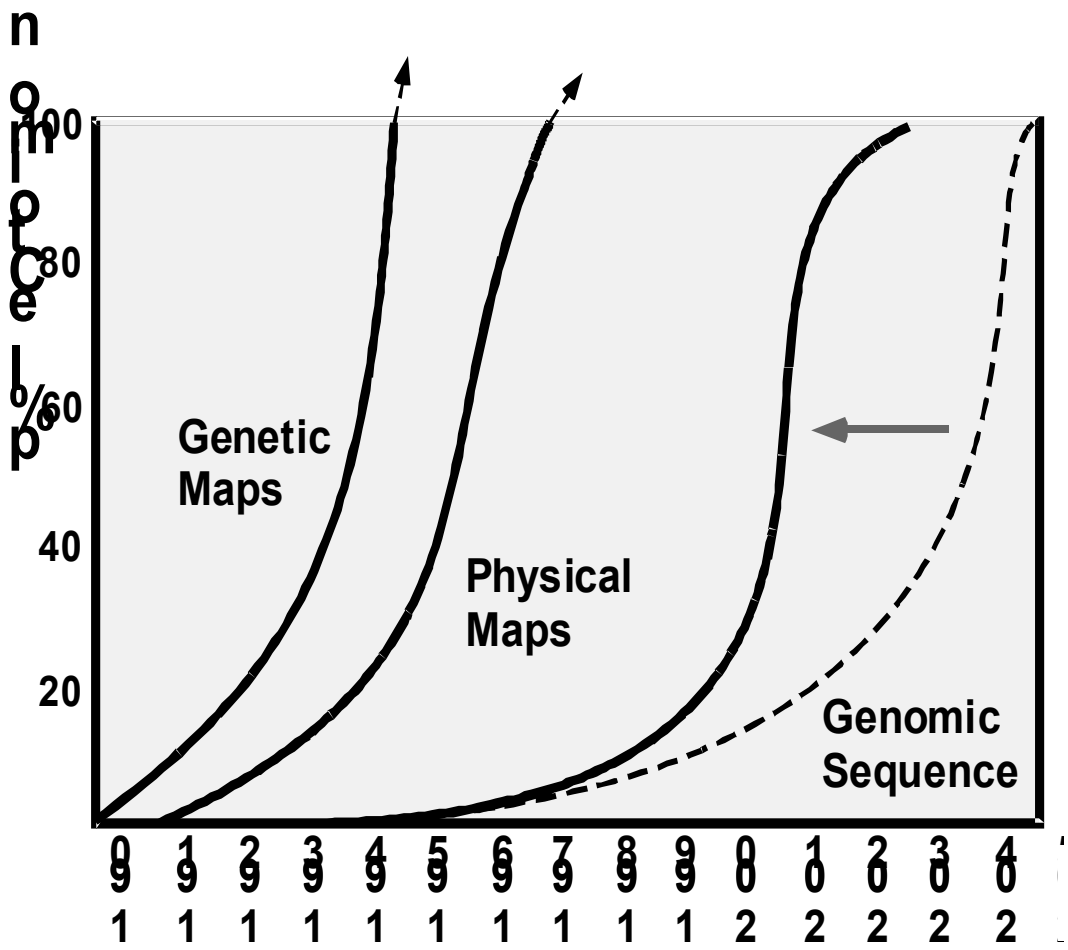
Human Genome Project: 5 Year Goals

New Goals for the U.S. Human Genome Project: 1998–2003

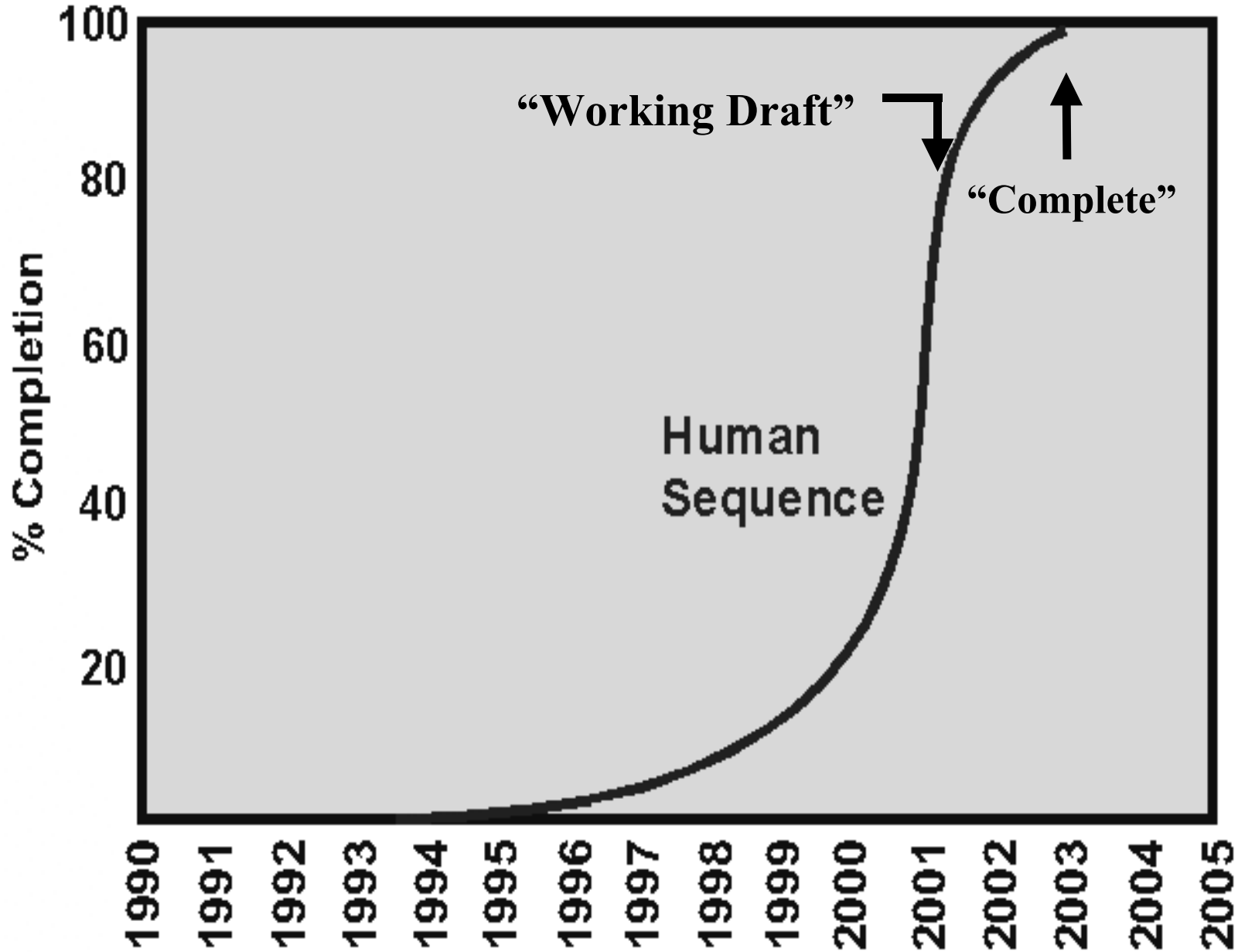
Francis S. Collins,* Ari Patrinos, Elke Jordan, Aravinda Chakravarti, Raymond Gesteland, LeRoy Walters, and the members of the DOE and NIH planning groups

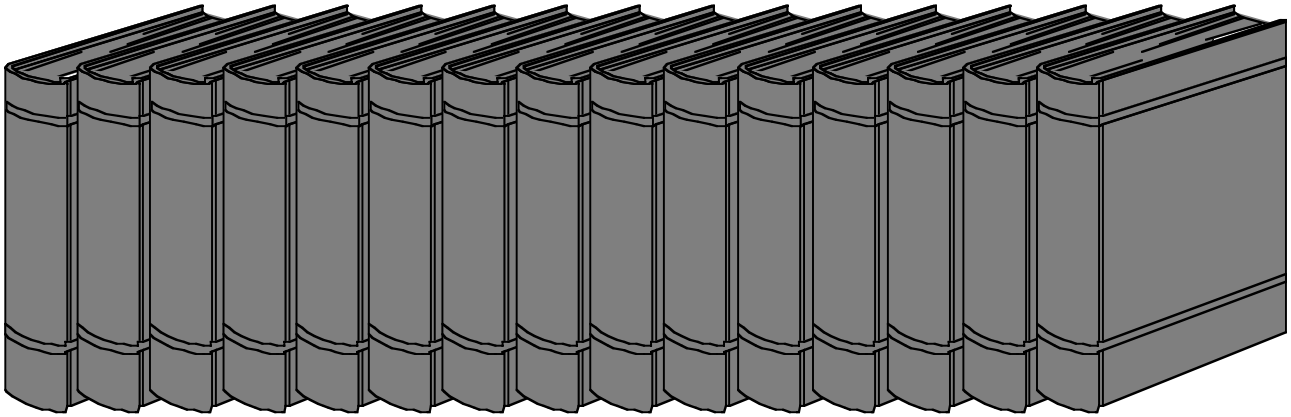
Science 282:682-689, 1998

Revised Timetable for Sequencing

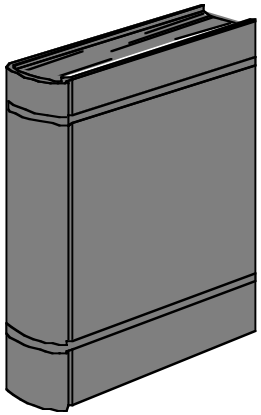


Timetable for Human Genome Sequencing





Genome
(~3000 Mb)



Chromosome
(~130 Mb)

GATCGTCTAGAATCTC
GAGATCTCTGAGAGTC
GTGGGAAACTGTGTGA
TGTGACTAGCCACAGT

TACGTGTGAGAGATGT
ATGATGCACCTGACCC
GGGTTTCACTCTCAAC
GACTCACTCCACCTCA

GAGGCCCACCGCCGCT
GTGCACGTCCACCACC

BAC
(~0.1-0.2 Mb)

“Sequence-Ready Maps”

Human Genome Sequencing Centers

Genome Sequencing Center
Washington University
School of Medicine
St. Louis, MO USA

Washington
WASHINGTON UNIVERSITY IN ST. LOUIS

Whitehead Institute/MIT Genome Sequencing Center

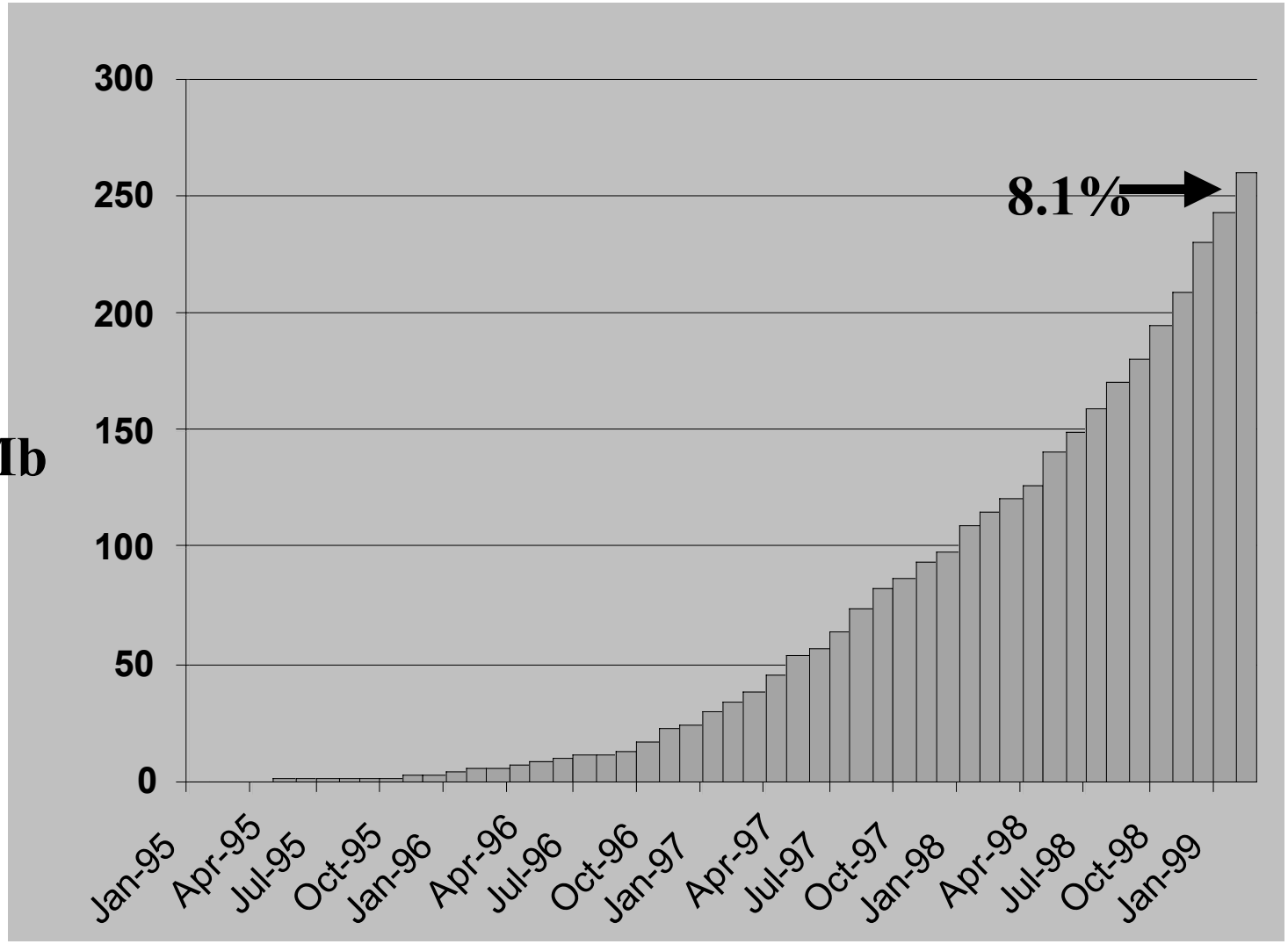
Baylor College of Medicin

HGSC
HUMAN GENOME SEQUENCING CENTER

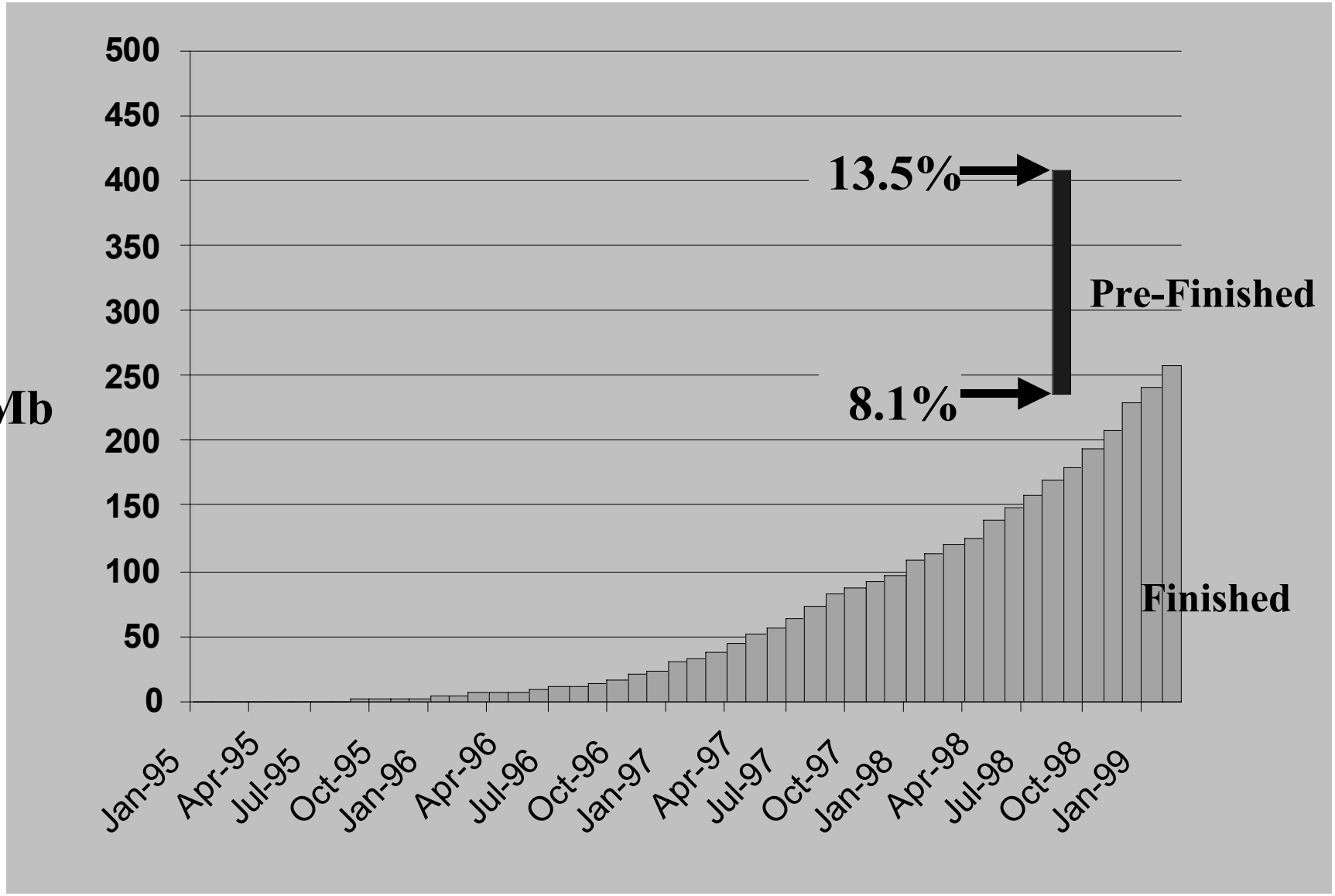
JGI
JOINT GENOME INSTITUTE

The Sanger Centre

Finished Human Genome Sequence



Total Human Genome Sequence



Human Genome Sequence by the HGP

- **Immediate Release**

  Sequence Contigs >1-2 kb

  Finished and Pre-Finished Sequence

- **High Accuracy**

  Error Rate of <1 in 10,000 bp

  Assessed/Confirmed by QC Exercises
  (see *Genome Research* 9:1-4, 1999)

- **Cost**

  Steady (But Not Massive) Decrease

  Currently at ~30-60¢ per Finished bp

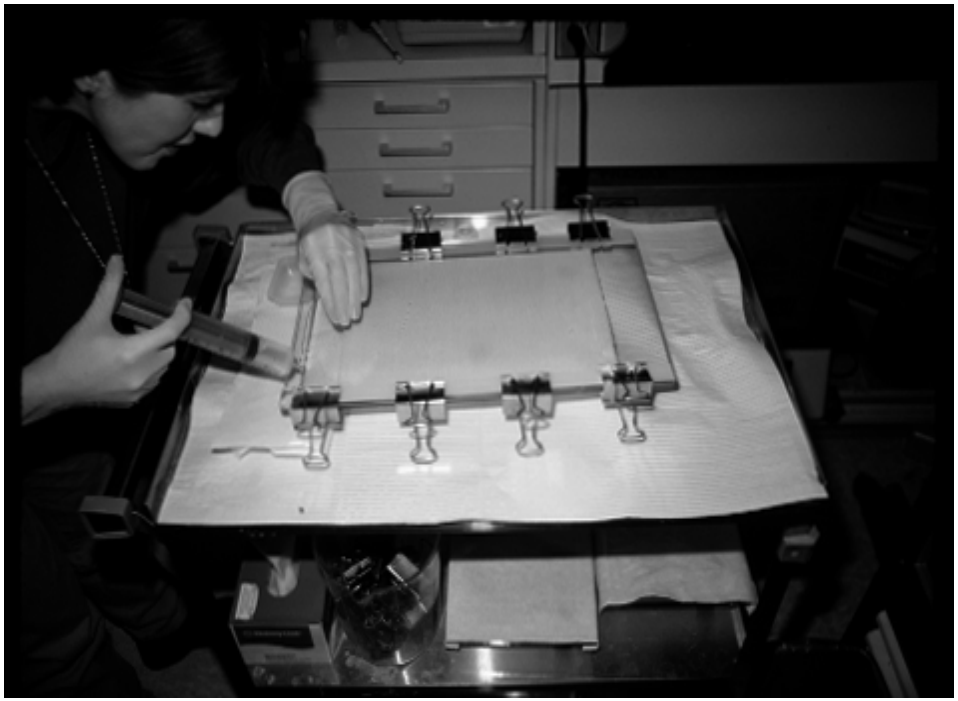
Issues for Scaling Genomic Sequencing

- **More Efficient Finishing Process**
- **Management**
- **Sample Processing (Automation)**
- **Sequencing Instruments**

Automation for Large-Scale Sequencing

Limitations of Gel-Based Systems

Gel Pouring

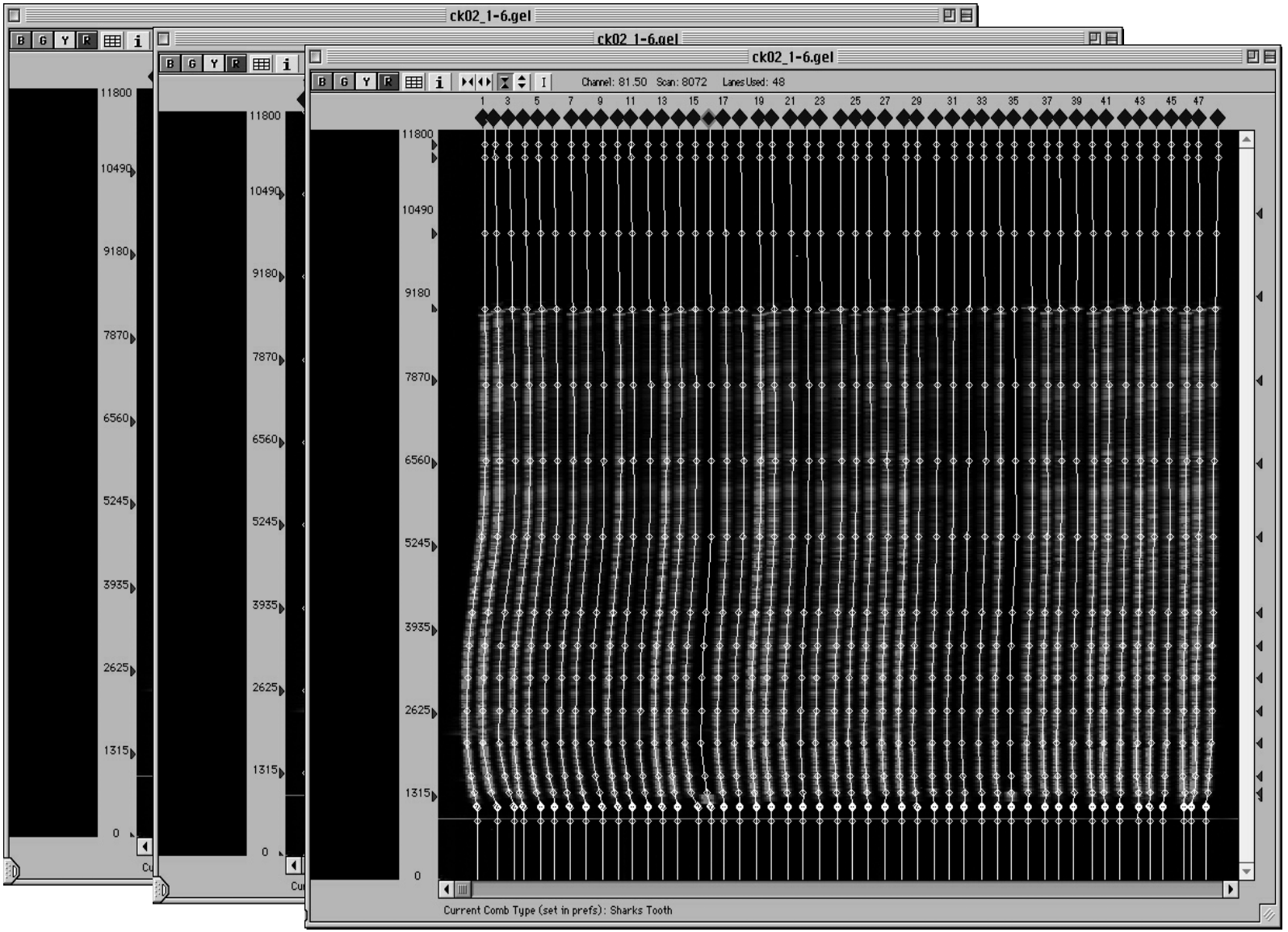


Gel Loading

Lots of Sequence Reads...

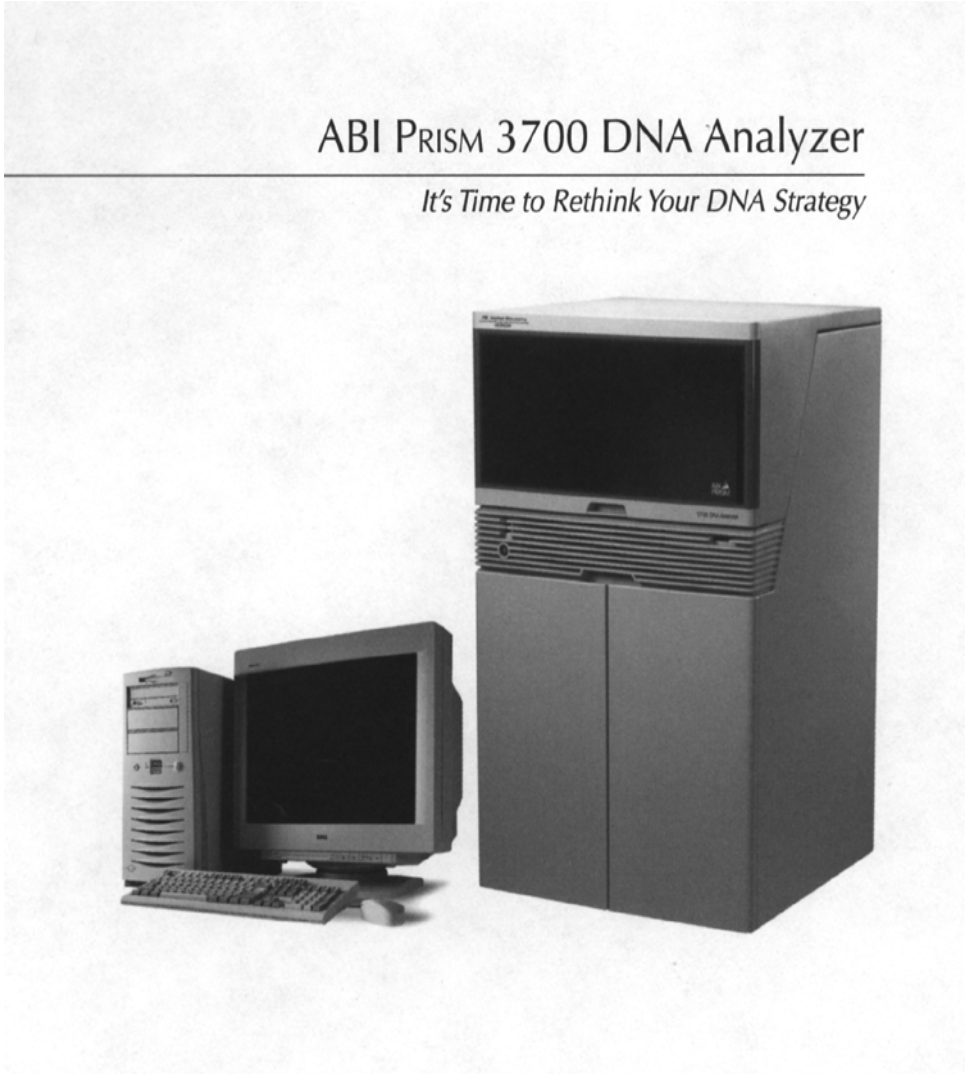
Lots of Gels!

Effort of Re-Tracking Gels

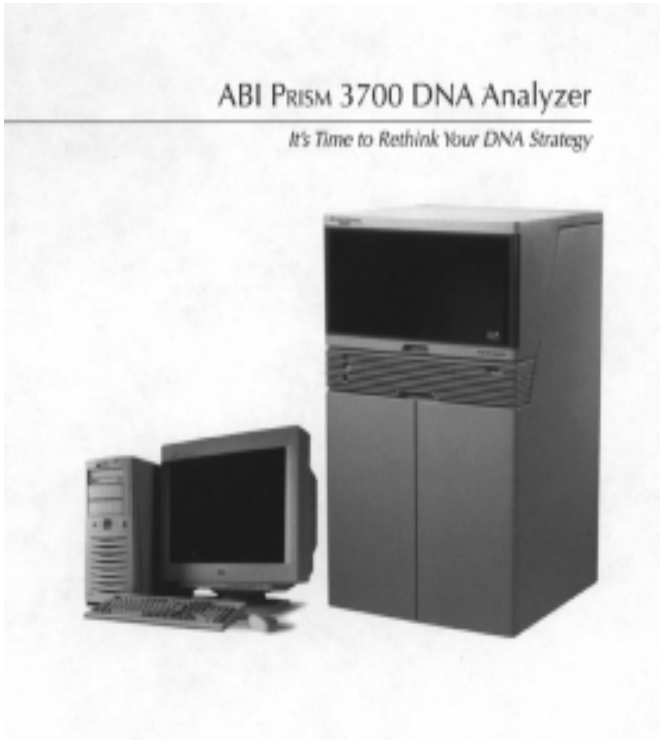


Capillary-based Sequencers

Applied Biosystems 3700



Commercial Interest in Human Genome Sequencing



CELERA

Whole-genome Shotgun Sequencing

Pros:
(Weber and Myers, 1997)

No sequence-ready maps required!
Savings of effort and cost

Much faster than clone-by-clone

Detection of DNA polymorphisms
Sequence 5 individuals

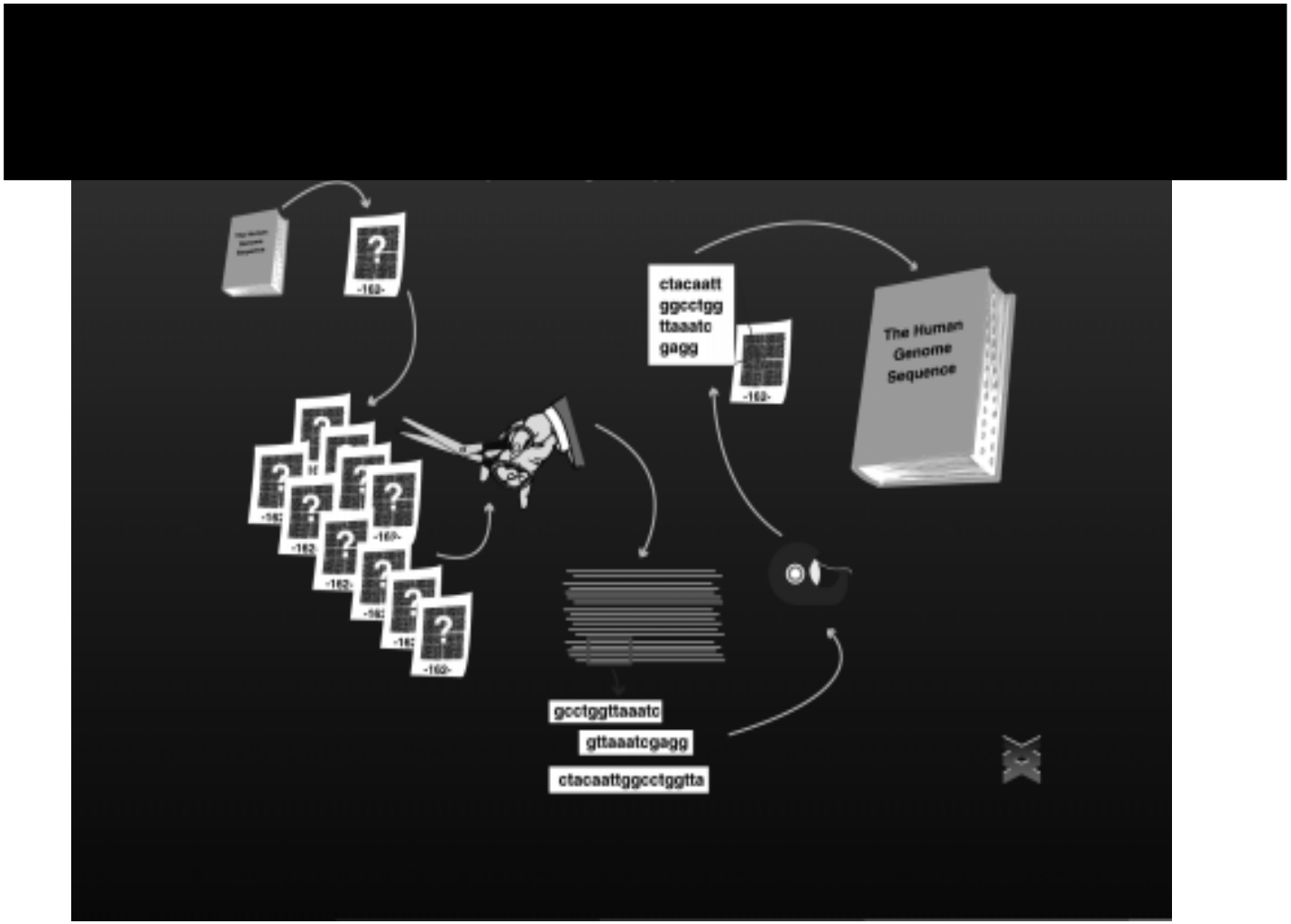
Cons:
(Green, 1997)

Probability of success debatable

Dubious cost savings
"Finishing" could be a mess

Final quality level contentious

Sequencing Mapped DNA



“Working Draft” Sequence

Skimming

0.5 -1X (~2 reads/kb)

Half-Shotgun

3-4X (~9 reads/kb)

Pre-Finished

6-8X (~18 reads/kb)

Full Shotgun

8-10X (~22 reads/kb)

Quality/Utility of "Working Draft" Sequence

GENOME METHODS

Analysis of the Quality and Utility of Random Shotgun Sequencing at Low Redundancies

John Bouck,[1,3] Webb Miller,[2] James H. Gorrell,[1] Donna Muzny,[1] and Richard A. Gibbs[1]

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030 USA;
[2]Department of Computer Science and Engineering, Pennsylvania State University, University Park, Pennsylvania 16802 USA

Genome Research **8:1074-1084, 1998**

Mouse Genome Analysis

An action plan for mouse genomics

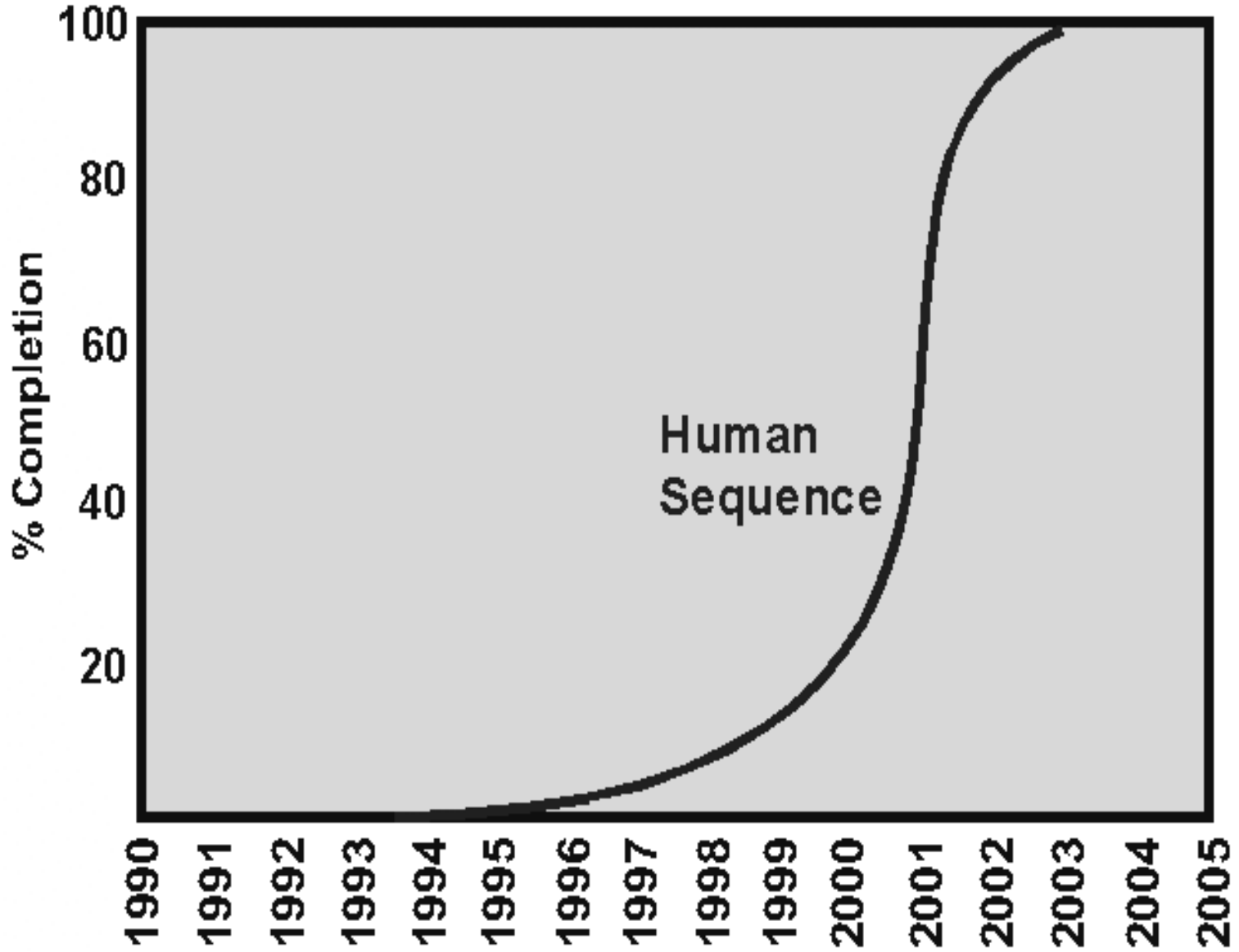
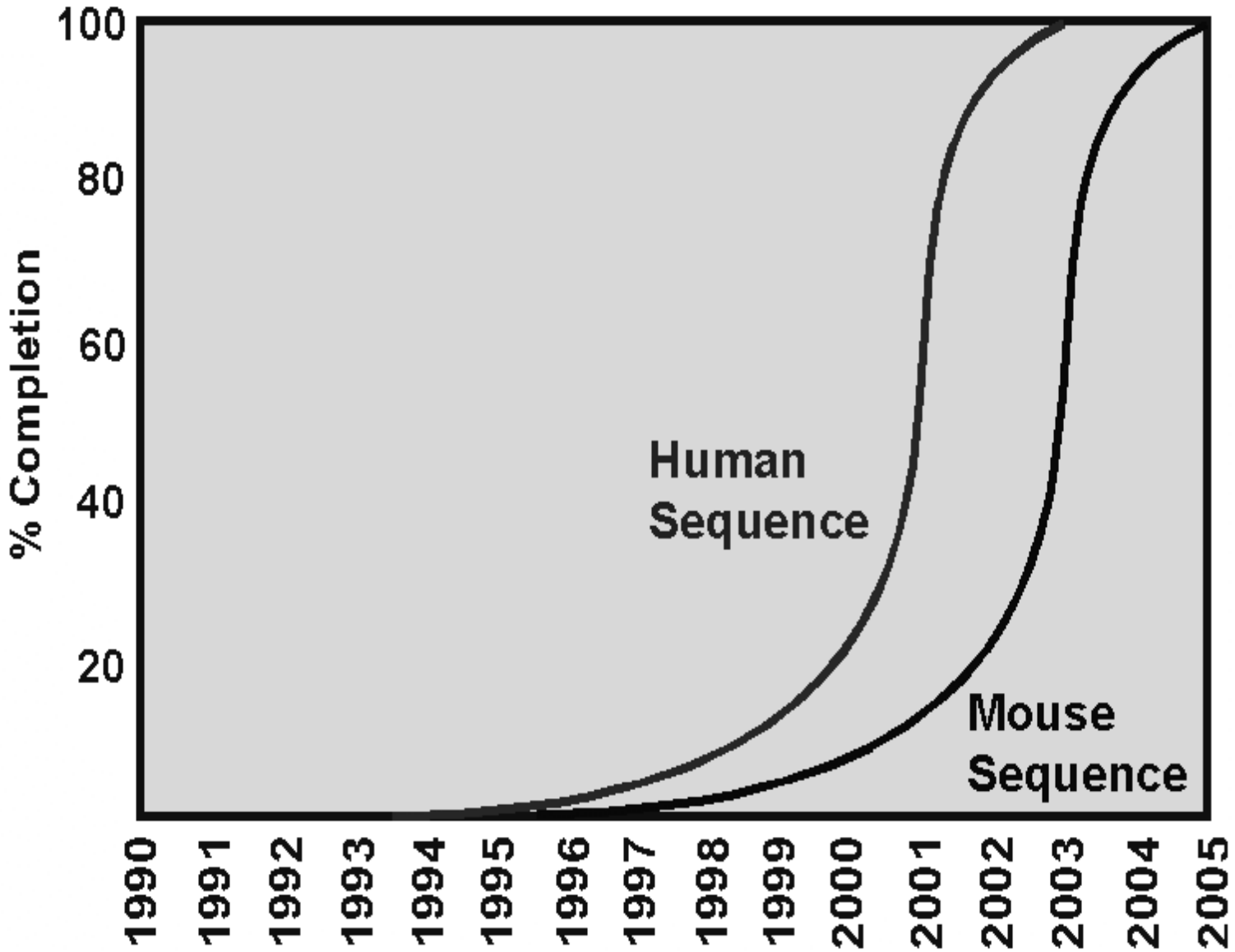
James Battey[1], Elke Jordan[2], David Cox[3] & William Dove[4]

The mouse has become the leading animal model for studying biological processes in mammals. Creation of additional genomic and genetic resources will make the mouse an even more useful model for the research community. On the basis of recommendations from the scientific community, the National Institutes of Health (NIH) plans to support grants to generate a 'working draft' sequence of the mouse genome by 2003, systematic mutagenesis and phenotyping centres, repositories for mouse strain maintenance, distribution and cryopreservation and training fellowships in mouse pathobiology.

http://www.nih.gov/science/mouse

Nature Genetics **21:73-75, 1999**

Highly Ambitious…



Even More Ambitious…